Supplementary material for "Template Inversion Attack against Face Recognition Systems using 3D Face Reconstruction"

Hatef Otroshi Shahreza[1,2] and Sébastien Marcel[1,3]
[1]Idiap Research Institute, Martigny, Switzerland
[2]École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[3]Université de Lausanne (UNIL), Lausanne, Switzerland
{hatef.otroshi,sebastien.marcel}@idiap.ch

Abstract

In this supplement, we first report the recognition performance of FR models used in the main paper in Sec. 1. Next, we provide the evaluation of blackbox and whitebox attacks against these SOTA FR models in Sec. 2 and Sec. 3, respectively. We also present samples of reconstructed faces (i.e., 3D reconstructed face, frontal reconstructed image, and reconstructed face image using the camera parameters grid). Finally, in Sec. 4 we report an ablation study on the hyperparameters of the camera parameter optimization for both our grid search (GS) and continuous optimization (CO) approaches. The project page is available at: https://www.idiap.ch/paper/gafar

1. Recognition Performance of FR models

In our experiments, we consider different SOTA FR models including ArcFace [2], ElasticFace [1] as well as four different FR models with SOTA backbones from FaceX-Zoo [10], including AttentionNet [9], HRNet [11], RepVGG [3] and Swin [6]. Tab. 1 reports the recognition performance of these FR models in terms of true match rate (TMR) at the thresholds correspond to false match rates (FMRs) of 10^{-2} and 10^{-3} evaluated on the LFW, MOBIO, and AgeDB datasets.

Table 1: Recognition performance of face recognition models used in our experiments in terms of true match rate (TMR) at the thresholds correspond to false match rates (FMRs) of 10^{-2} and 10^{-3} evaluated on the LFW, MOBIO, and AgeDB datasets. The values are in percentage.

model	**LFW**		**MOBIO**		**AgeDB**	
	FMR=10^{-2}	**FMR=10^{-3}**	**FMR=10^{-2}**	**FMR=10^{-3}**	**FMR=10^{-2}**	**FMR=10^{-3}**
ArcFace	97.60	96.40	100.00	99.98	98.33	98.07
ElasticFace	96.87	94.70	100.00	100.00	98.20	97.57
AttentionNet	84.27	72.77	99.71	97.73	97.93	96.90
HRNet	89.30	78.43	98.98	98.23	97.67	96.23
RepVGG	77.20	58.07	98.75	95.80	95.93	93.93
Swin	91.70	87.83	99.75	98.98	98.03	97.10

2. Evaluation of *blackbox* attack against SOTA FR models

To evaluate the proposed method in the *blackbox* attack against SOTA FR models, we consider ArcFace and also ElasticFace as F_{loss} to calculate the ID loss. Tab. 2 compares the performance of the proposed method with *blackbox* face reconstruction methods in the literature in terms of SAR for systems configured at false match rate (FMR) of 10^{-2}. Similar results for FMR $= 10^{-3}$ are reported in the main paper. Tab. 3 also reports similar evaluation for FMR $= 10^{-2}$ and FMR $= 10^{-3}$ on the AgeDB dataset. As these tables show the proposed method achieves superior performance than previous methods in the literature. Furthermore, camera parameter optimization (GS) improves the performance of GaFaR.

Table 2: Evaluation of *blackbox* attack against SOTA FR models at systems' FMR=10^{-2} on the LFW and MOBIO datasets in terms of success attack rate (SAR). For attacks using our proposed method, we use ArcFace and also ElasticFace as F_{loss} to calculate the ID loss. The values are in percentage.

	LFW						**MOBIO**					
	ArcFace	**ElsFace**	**Att.Net**	**HRNet**	**RepVGG**	**Swin**	**ArcFace**	**Els.Face**	**Att.Net**	**HRNet**	**RepVGG**	**Swin**
NBNetA-M [7]	14.30	37.13	10.37	20.19	10.64	13.18	2.85	10.00	4.76	4.76	6.19	6.67
NBNetA-P [7]	35.61	60.05	6.80	16.83	26.43	25.92	23.81	60.96	15.24	14.29	44.76	30.48
NBNetB-M [7]	26.90	52.99	17.62	31.74	18.17	27.00	20.95	30.00	21.43	25.24	21.43	27.62
NBNetB-P [7]	61.66	81.74	43.41	56.30	38.12	61.02	49.05	70.95	0	64.76	51.43	71.43
Dong *et al.* [4]	28.21	34.56	19.17	24.87	14.76	26.62	24.29	34.76	38.57	16.19	24.76	18.10
Vendrow and Vendrow [8]	77.00	79.37	46.52	49.52	22.40	66.07	69.52	74.29	55.71	43.81	39.52	70.00
[F_{loss}= Els.Face] GaFaR	71.25	-	39.11	35.83	28.64	63.40	77.14	-	80.47	68.10	72.86	90.48
[F_{loss}= Els.Face] GaFaR + GS	**78.12**	-	46.79	45.45	35.25	68.92	**85.71**	-	86.19	78.57	77.61	91.90
[F_{loss}= ArcFace] GaFaR	-	89.78	56.57	67.64	46.89	78.91	-	91.90	89.05	87.62	87.14	96.19
[F_{loss}= ArcFace] GaFaR + GS	-	**91.28**	**62.03**	**72.28**	**51.27**	**81.39**	-	**93.33**	**90.00**	**90.00**	**90.95**	**96.19**

Table 3: Evaluation of *blackbox* attack against SOTA FR models at systems' FMR=10^{-2} and FMR=10^{-3} on the AgeDB dataset in terms of success attack rate (SAR). For attacks using our proposed method, we use ArcFace and also ElasticFace as F_{loss} to calculate the ID loss. The values are in percentage.

	FMR=10^{-2}						**FMR=10^{-3}**					
	ArcFace	**ElsFace**	**Att.Net**	**HRNet**	**RepVGG**	**Swin**	**ArcFace**	**ElsFace**	**Att.Net**	**HRNet**	**RepVGG**	**Swin**
NBNetA-M [32]	2.56	8.44	1.85	2.45	2.85	18.29	0.81	2.55	0.22	0.38	0.44	0.27
NBNetA-P [32]	9.30	20.07	2.42	1.54	10.14	4.72	3.99	8.92	0.34	0.14	3.71	1.02
NBNetB-M [32]	5.40	14.56	3.83	3.68	4.71	3.70	1.88	6.27	0.50	0.77	1.06	0.68
NBNetB-P [32]	23.89	44.46	17.19	14.83	18.62	21.48	13.18	28.94	5.08	5.61	7.92	8.75
Dong *et al.* [13]	9.13	12.10	7.58	6.01	6.81	7.62	3.93	4.88	1.58	1.97	2.22	2.48
Vendrow and Vendrow [47]	**44.74**	52.17	35.47	24.65	27.39	40.43	**29.64**	34.89	15.06	12.02	14.49	21.10
[F_{loss}= Els.Face] GaFaR	36.00	-	18.30	8.55	15.73	29.89	21.67	-	6.70	3.10	7.10	14.67
[F_{loss}= Els.Face] GaFaR + GS	44.37	-	26.03	14.07	21.98	36.65	28.94	-	9.70	6.05	11.46	19.62
[F_{loss}= ArcFace] GaFaR	-	63.45	33.23	31.56	31.71	49.17	-	47.37	14.59	17.09	18.02	30.05
[F_{loss}= ArcFace] GaFaR + GS	-	**68.82**	**40.26**	**38.53**	**38.78**	**55.20**	-	**53.10**	**18.76**	**22.40**	**24.01**	**35.20**

Table 4: Evaluation of *whitebox* attack against SOTA FR models at systems' FMR $= 10^{-2}$ on the LFW and MOBIO datasets in terms of success attack rate (SAR). The values are in percentage.

	LFW						**MOBIO**					
	ArcFace	**ElsFace**	**Att.Net**	**HRNet**	**RepVGG**	**Swin**	**ArcFace**	**Els.Face**	**Att.Net**	**HRNet**	**RepVGG**	**Swin**
GaFaR	89.27	84.25	49.05	61.62	39.22	83.56	95.71	90.0	87.62	85.24	78.57	97.14
GaFaR + GS	90.77	86.66	55.35	66.96	41.70	85.10	97.62	91.90	90.00	90.48	82.38	97.62
GaFaR + CO	**91.87**	**88.27**	**57.51**	**68.64**	**41.63**	**85.85**	**97.62**	**93.81**	**89.52**	**90.48**	**84.76**	**98.10**

3. Evaluation of *whitebox* attack against SOTA FR models

Tab. 4 reports the performance of the proposed method in the *whitebox* attack in terms of SAR for FR systems configured at false match rate (FMR) of 10^{-2}. Similar results for FMR $= 10^{-3}$ are reported in the main paper. **??** also reports similar evaluation for FMR $= 10^{-2}$ and FMR $= 10^{-3}$ on the AgeDB dataset. According to these tables, all these FR models are vulnerable to our attack. Furthermore, the camera parameter optimization (GS and CO) improves the performance of GaFaR. Comparing grid search with continuous optimization, results show that with the same number of iterations continuous optimization achieves better performance. Figs. 1 to 4 illustrate sample face images from the FFHQ dataset and their reconstructed faces from ArcFace templates in the *whitebox* attack. These figures present the 3D reconstructed faces, frontal reconstructed images, and the grid of reconstructed face images with different camera hyperparameters (i.e., grid size, interval of Φ, and interval of Θ) used in our experiments for our grid search optimization approach in the main paper.

4. Ablation Study on Camera Parameter Optimization

In this section, we report an ablation study on the hyperparameters of the camera parameter optimization for both our grid search (GS) and continuous optimization (CO) approaches.

Table 5: Evaluation of *whitebox* attack against SOTA FR models at systems' FMR $= 10^{-2}$ and FMR $= 10^{-3}$ on the AgeDB dataset in terms of success attack rate (SAR). The values are in percentage.

	FMR=10^{-2}						**FMR=10^{-3}**					
	ArcFace	**ElsFace**	**Att.Net**	**HRNet**	**RepVGG**	**Swin**	**ArcFace**	**Els.Face**	**Att.Net**	**HRNet**	**RepVGG**	**Swin**
GaFaR	63.30	53.54	23.60	23.29	13.29	59.90	48.94	36.72	9.13	12.00	5.81	39.94
GaFaR + GS	67.86	60.90	30.51	27.32	17.94	63.20	53.10	43.10	12.64	14.21	9.13	41.94
GaFaR + CO	**71.95**	**63.63**	**34.68**	**32.31**	**19.31**	**67.08**	**58.00**	**47.22**	**14.87**	**18.35**	**9.99**	**47.72**

Grid Search (GS) For grid search, in the main paper, we consider $\psi \in [-45°, +45°]$ and $\theta \in [-30°, +30°]$ for a 11×11 grid with step sizes of $\psi_{\text{step}} = 9°$ and $\theta_{\text{step}} = 6°$. In this ablation study, we consider whitebox attack against ArcFace and in each experiment, we change one of these hyperparamters (i.e., grid size, interval of Φ, and interval of Θ) and evaluate the effect in terms of SAR and average execution time. Fig. 5 reports our ablation study for a system configured at FMR=10^{-3} on the MOBIO dataset. According to this ablation study, the intervals of Φ and Θ should not be necessarily very large. Furthermore, by increasing the size of the grid we can have higher SAR with the cost of more execution time.

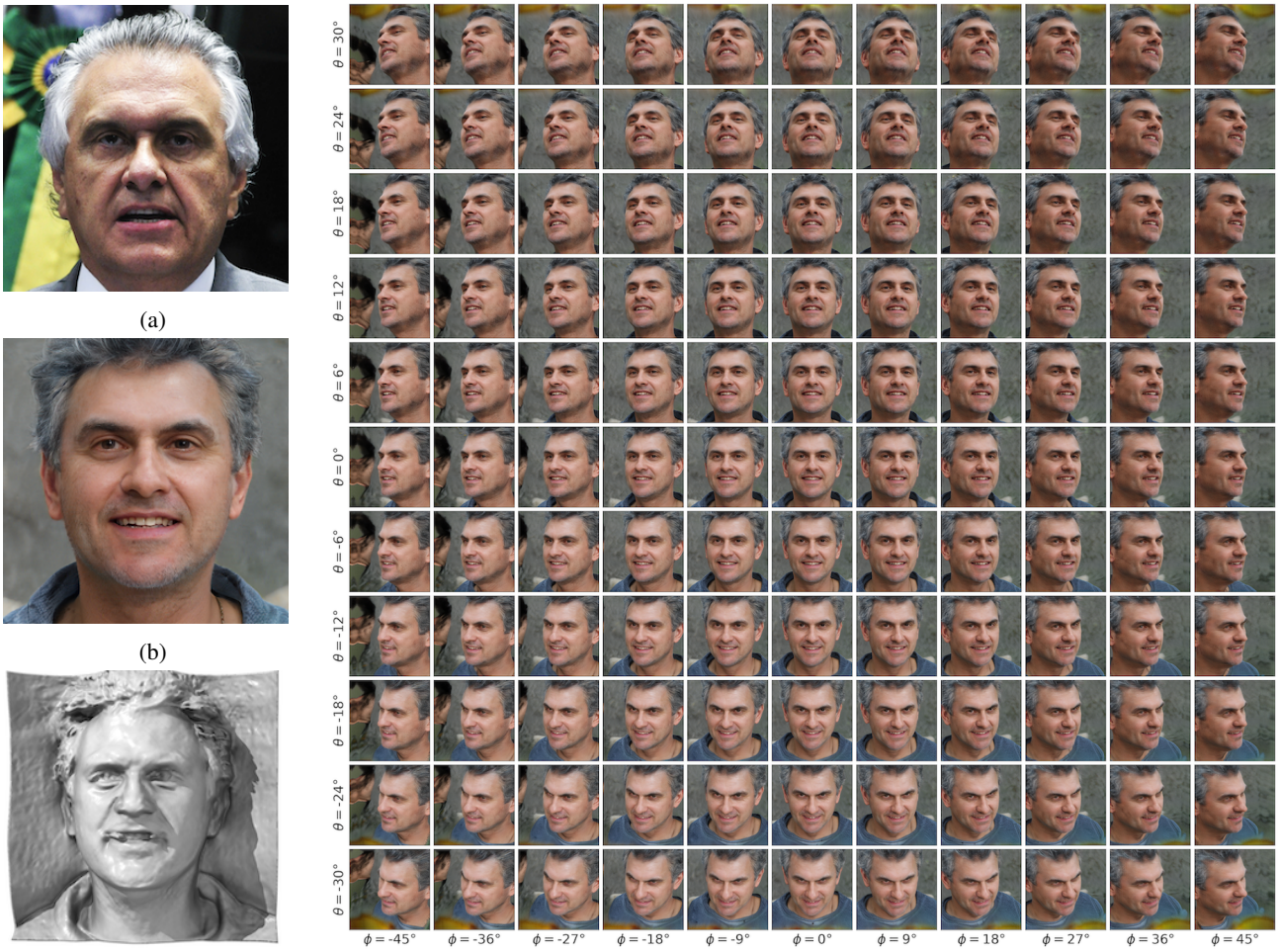
Continuous Optimization (CO) In the continuous optimization approach, we consider $\psi \in [-45°, +45°]$ and $\theta \in [-30°, +30°]$ and use 121 iterations of Adam optimizer [5] with the learning rate of 10^{-2} in the main paper. In this ablation study, we consider whitebox attack against ArcFace and in each experiment, we change one of these hyperparamters (i.e., learning rate, number of iterations, interval of Φ, and interval of Θ) and evaluate the effect in terms of SAR and average execution time. Fig. 6 reports our ablation study for a system configured at FMR=10^{-3} on the MOBIO dataset. Based on these results, the intervals of Φ and Θ should not be necessarily very large similar to ablation study for the grid search optimization. Moreover, by increasing the number of iterations we can have higher SAR with the cost of more execution time.

(a) (b) (c) (d)

Figure 1: Sample face image from the FFHQ dataset (a), its frontal reconstructed face image (b), its 3D face reconstruction (c), and the corresponding reconstructed face images with camera parameters grid (d) using our method in the *whitebox* attack against ArcFace. The cosine similarity between templates of original (a) and frontal (b) reconstructed face images is 0.738.

(a) (b) (c) (d)

Figure 2: Sample face image from the FFHQ dataset (a), its frontal reconstructed face image (b), its 3D face reconstruction (c), and the corresponding reconstructed face images with camera parameters grid (d) using our method in the *whitebox* attack against ArcFace. The cosine similarity between templates of original (a) and frontal (b) reconstructed face images is 0.681.



(a) (b) (c) (d)

Figure 3: Sample face image from the FFHQ dataset (a), its frontal reconstructed face image (b), its 3D face reconstruction (c), and the corresponding reconstructed face images with camera parameters grid (d) using our method in the *whitebox* attack against ArcFace. The cosine similarity between templates of original (a) and frontal (b) reconstructed face images is 0.677.

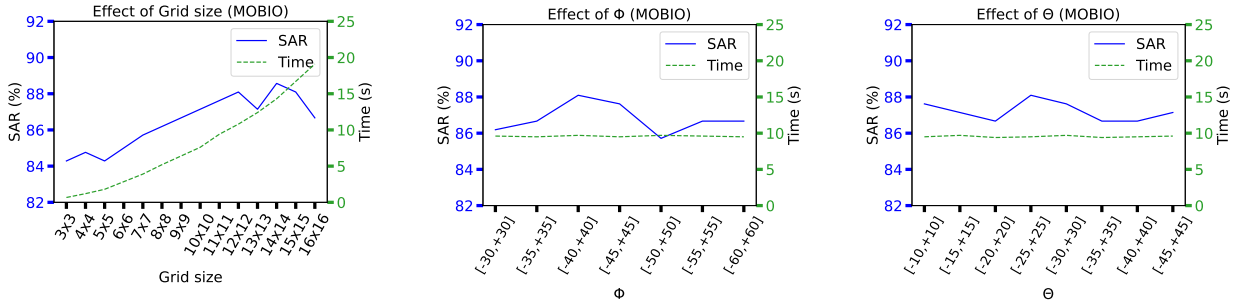
(a)

(b)

(c)

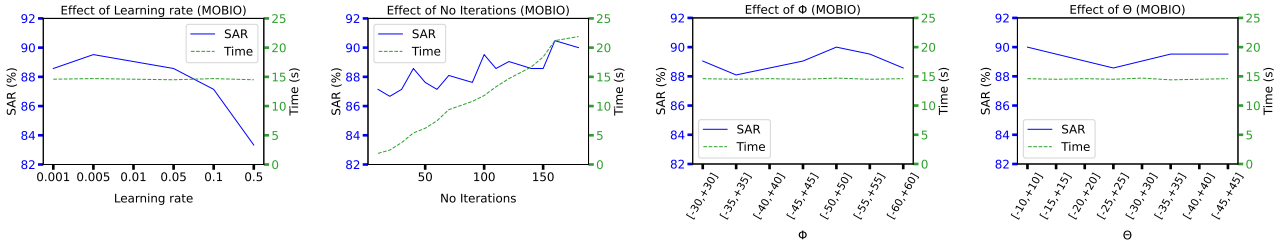
(d)

Figure 4: Sample face image from the FFHQ dataset (a), its frontal reconstructed face image (b), its 3D face reconstruction (c), and the corresponding reconstructed face images with camera parameters grid (d) using our method in the *whitebox* attack against ArcFace. The cosine similarity between templates of original (a) and frontal (b) reconstructed face images is 0.679.



(a) (b) (c)

Figure 5: Ablation study on the effect of different hyperparamters in grid search for camera parameters optimization in terms of success attack rate (SAR) and average execution time for each image reconstruction for whitebox attack ArcFace configured at FMR=10^{-3} on the MOBIO dataset: a) grid size, b) interval of Φ, and c) interval of Θ.



(a) (b) (c) (d)

Figure 6: Ablation study on the effect of different hyperparamters in continuous optimization for camera parameters in terms of success attack rate (SAR) and average execution time for each image reconstruction for whitebox attack ArcFace configured at FMR=10^{-3} on the MOBIO dataset: a) learning rate, b) number of iterations, c) interval of Φ, and d) interval of Θ.

References

[1] Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Elasticface: Elastic margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1578–1587, 2022. 1

[2] Jiankang Deng, Jia Guo, Xue Niannan, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1

[3] Xiaohan Ding, Xiangyu Zhang, Ningning Ma, Jungong Han, Guiguang Ding, and Jian Sun. Repvgg: Making vgg-style convnets great again. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13733–13742, 2021. 1

[4] Xingbo Dong, Zhe Jin, Zhenhua Guo, and Andrew Beng Jin Teoh. Towards generating high definition face images from deep templates. In *2021 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–11. IEEE, 2021. 2

[5] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of the International Conference on Learning Representations (ICLR)*, San Diego, California., USA, May 2015. 3

[6] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (CVPR)*, pages 10012–10022, 2021. 1

[7] Guangcan Mai, Kai Cao, Pong C Yuen, and Anil K Jain. On the reconstruction of face images from deep face templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(5):1188–1202, 2018. 2

[8] Edward Vendrow and Joshua Vendrow. Realistic face reconstruction from deep embeddings. In *NeurIPS 2021 Workshop Privacy in Machine Learning*, 2021. 2

[9] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. Residual attention network for image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3156–3164, 2017. 1

[10] Jun Wang, Yinglu Liu, Yibo Hu, Hailin Shi, and Tao Mei. Facex-zoo: A pytorch toolbox for face recognition. In *Proceedings of the 29th ACM International Conference on Multimedia*, 2021. 1

[11] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 1